

数据挖掘在生物医学数据分析中的应用研究

莫雨

河南大学，河南

摘要

在生物医学的研究中，为了提升生物医学研究的精确性，需要对生物医学研究的数据进行采集、存储和管理工作，我国现今的数据分析方案需要进行完善，因此在生物医学的数据分析中，数据挖掘这门学科便逐渐发展起来了。本文主要对数据挖掘的概念进行阐述，在将数据挖掘应用到生物医学中进行研究分析。

关键词：数据挖掘；生物医学；统计学

ABSTRACT

In biomedical research, in order to improve the accuracy of biomedical research, data collection, storage and management of biomedical research are needed. Today's data analysis programs in China need to be improved. Therefore, in biomedical data analysis, The discipline of data mining has gradually developed. This paper mainly expounds the concept of data mining, and applies data mining to biomedicine for research and analysis.

Key words: data mining; biomedicine; statistics

引言

所谓数据挖掘，数据挖掘又称职位数据库的调查发现。数据挖掘实验主要针对的在生物医学实验中对大量的实验数据进行研究分析，并从实验数据中找到数据之间前后的关联性。随着我国信息技术的不断发展，在生物医学的实验中，对实验数据的采集、存储和管理的方式。随着技术的发展，数据分析也更加具有精确性，因此在生物医学的研究分析中，针对大量的数据分析中如果只是通过传统的分析方法进行分析，只能获取实验数据中表面信息，不能数据信息进行深入分析。因此在数



<http://ijmd.oajrc.org>

OPEN ACCESS

DOI: 10.12208/ijmd.20190001

收稿日期: 2019-06-04

出刊日期: 2019-06-25

莫雨

河南大学，河南

据分析中,数据挖掘在这一环境下便运营而生了。本文主要针对数据挖掘在生物医学的研究进行讨论分析。

1、生物医学研究中数据挖掘的应用

数据挖掘在生物医学中的运用,最常见的数据分析方法主要有以下几种,下列将具体作出阐述分析。

1.1 生物医学中的预测型数据挖掘

1.1.1 数据挖掘的分类

在数据挖掘中分类,主要是针对生物医学数据实验类别进行分类研究托西,在数据挖掘中分类,需要建立数据分类模型,这种分类模型也称之为分类器,分类器主要针对实验数据的具体特征和数据与数据之间的关联性进行分类处理。并数据的类别收集于分类器中,在生物医学实验中数据挖掘常用的方法主要有,1 决策树法、2 贝叶斯分类法等这两种基本方法。在生物医学实验研究的应用中,对疾病的诊断和疾病鉴别,我们可知在生物医学的研究中,诊断也便是一个数据分类过程。在数据研究分析的运用中,通过美国麻省理工学院,所建立的心律失常数据库的数据模型进行研究分析,这种数据模型的建立通常是采用 1SVM 分类器、2KNN 分类器和 3ANN 这三种数据模型建立,这种数据模型能够在数据分析中对心电图常见异常波形和正常波形进行数据分类处理。通过调查实验发现这三种数据分析处理器的正确率分别为 85%、83% 和 82%,实验表明了数据分析的精确性对生物医学研究的重要性。

1.1.2 挖掘数据的回归分析

在挖掘数据的分析中,回归分析通常是指生物医学实验中的实验数据的自变量,在数据

自变量的设置时,可以选取多个自变量(多元回归),数据自变量的设置时,是通过数据分析建立数据模型。因此在数据回归模型建设中,需要保证数据模型建立的准确性,在生物医学实验分析时,最常用的数据回归分析,主要有 1 线性回归、2Logistic 回归、3ANN、回归树。这三种基本的数据分析,由此可见数据回归分析,在生物医学研究中运用的广泛性。

挖掘数据回归分析在生物医学的研究应用中,还包括生物医疗诊断和生物医疗判别和多种病症的研究。在数据的研究分析中,传统的数据分析方法对实验数据进行分析,挖掘数据的回归分析比传统数据分析方法更加具有精确性,而回归数据分析主要优势在于能够对数据进行针对性分析。

2、生物医学研究中描述型数据挖掘

2.1 数据的关联性分析

在挖掘数据分析中,数据的关联性分析主要分析数据之间是否具有一定的关联性,这种数据分析法,能够真实的检测数据与数据之间的关联性。可以科学的预测某类事件发生的可能性。在生物医学数据分析中,数据之间的关联性并非只有一种因果关系。而在生物医学的研究领域中,数据也存在数据因果性。比如在数据分析中经常发现,某种疾病能够表现出不同的多种症状,这些症状之间也具有一定的关联性。由于临床医学上的鉴别,是建立在症状之间的检测方法,总之在生物医学的数据研究中,数据挖掘分析的方法有很多,但传统的研究方法具有一定的局限性,我们需要将这种分析方法进行完善。

3、数据挖掘在生物医学中前景展望

通过上述的研究分析，生物医学的数据分析众多。但在这些数据分析方法在实际的医学运用中还处于起步阶段，其中主要的原因有以下两个方面。

(1) 在数据挖掘中，需要运用到大量复杂的数学工具，因此在生物医学的研究中出现“望而却步”的局面。其实，这是挖掘数据发展“恐惧”心理，这种“恐怖”心理是不必要的。但在生物医学研究分析中，对数学分析工具的运用性较少，在数据分析中，需要运用到大量的数据处理软件。（如 Weka、DBMiner、SPSS Clementine 和 SAS Enterprise Miner 等）这些软件中包含了大量的数据处理方法。

(2) 在我国现今的生气医学的“数据挖

掘”分析中，很难针对具体情况而进行选择数据分析方法。针对数据分析的这种局面，有些大学和医院在研究生课程中开设了相关专业课程，比如说在上海交通大学医学院也开设了《生物医学数据挖掘》的课程，以使生物医学相关工作认为能够理性认识数据挖掘分析。

4、总结

通过上述讨论分析，我们认识到数据挖掘在生物医学应用研究中的重要性。随着数据挖掘技术的发展，生物医学相关工作人员对“数据挖掘”的应用理解也不断加深，这种新型的数据分析方法对我国生物医学科技的发展有着促进作用。

参考文献

- [1] Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery[M]//Advances in data mining and knowledge discovery. Menlo Park: AAAI Press, 2017:1-34.
- [2] Han JW, Kamber M. Data mining: concepts and techniques[M]. 2nd ed. San Francisco: Morgan Kaufmann, 2018.
- [3] Tan PN, Steinbach M, Kumar V. Introduction to data mining[M]. 北京: 人民邮电出版社, 2018.
- [4] Benson DA, Karsch-Mizrachi I, Lipman DJ, et al. GenBank. Nucleic acids research[DB/OL]. 2018. <http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html>.